Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Computational and Applied Mathematics 180 (2005) 201–211

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Global convergence on an active set SQP for inequality constrained optimization☆

Xin-Wei Liu[a,b,*]

[a] Department of Applied Mathematics, Mathematics, Hebei University of Technology, Tianjin, China
[b] Singapore-MIT Alliance, National University of Singapore, E4-04-10, 4 Engineering Drive 3, Singapore 117576, Singapore

Received 28 November 2003

Abstract

Sequential quadratic programming (SQP) has been one of the most important methods for solving nonlinearly constrained optimization problems. In this paper, we present and study an active set SQP algorithm for inequality constrained optimization. The active set technique is introduced which results in the size reduction of quadratic programming (QP) subproblems. The algorithm is proved to be globally convergent. Thus, the results show that the global convergence of SQP is still guaranteed by deleting some "redundant" constraints.

Keywords: Active set; Sequential quadratic programming; Nonlinearly constrained optimization

1. Introduction

Sequential quadratic programming (SQP) has been one of the most important methods for solving nonlinearly constrained optimization problems. It is well known that SQP is very efficient for solving medium and small size nonlinear programs (see [1,4,6,23]). Recently, SQP has been applied to some problems such as nonsmooth equations, variational inequality problems, mathematical programs with equilibrium constraints (MPEC), etc. (see [5,8,15,22]), and some large-scale problems (see [13,14]). All

☆ Research is partially supported by SRF for ROCS, SEM and a doctoral fund from Hebei province, China.
* Corresponding address. Department of Applied Mathematics, Hebei University of Technology, Tianjin 300130, China. Tel.: +86 22 26564339; fax: +86 22 26564469.

E-mail address: mathlxw@jsmail.hebut.edu.cn (X.-W. Liu).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2004.10.012

these works show that not only is SQP still a very active topic in the research of numerical optimization, but also very useful approach in solving optimization problems.

Our motivation for this work is originated from the applications of SQP in solving large-scale problems. We consider the standard inequality constrained optimization problem

$$\min \quad f(x) \tag{1.1}$$

$$\text{s.t.} \quad c(x) \geqslant 0, \tag{1.2}$$

where $f(x): \Re^n \to \Re^1$, $c(x): \Re^n \to \Re^m$ are twice continuously differentiable functions. SQP is a kind of iterative method, at iterate k it needs to solve a QP subproblem

$$\min \quad g_k^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} B_k d \tag{1.3}$$

$$\text{s.t.} \quad c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d \geqslant 0, \quad i = 1, \ldots, m, \tag{1.4}$$

where x_k is the current iterate, $g_k = \nabla f(x_k)$ and B_k is an $n \times n$ matrix. Often B_k is required to be positive definite, which is supposed to be an approximate Hessian of the Lagrangian

$$\ell(x, \lambda) = f(x) - \lambda^{\mathrm{T}} c(x) \tag{1.5}$$

at the current iterate x_k. A merit function, which is normally a penalty function such as the ℓ_1 exact penalty function, is used to carry out line searches. It has been proved that SQP is globally convergent (see [7,16,17,21]).

One of the resulted difficulties of SQP for large-scale problem is that the memory requisite for each QP subproblem may be very large if the original problem is large (with great number of constraints). Our idea is originated from the following observations. Let x^* be a local solution of the original problem (1.1)–(1.2), the active set at x^* is defined by

$$A(x^*) = \{i \mid c_i(x^*) = 0\}. \tag{1.6}$$

It is easy to see that x^* is a local solution of (1.1) subject to

$$c_i(x) \geqslant 0, \quad i \in A(x^*), \tag{1.7}$$

or subject to

$$c_i(x) = 0, \quad i \in A(x^*). \tag{1.8}$$

Moreover, for many problems, the number of active constraints is much smaller than the total number of constraints. Therefore, problem (1.1) with (1.7) or problem (1.1) with (1.8) has less constraints than the original problem (1.1)–(1.2). For the QP subproblem (1.3)–(1.4), we can use active set techniques to subproblems with fewer constraints. For example, in the REQP method of [2], the constrained conditions in the QP subproblem is

$$\hat{c}_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d = 0, \quad i \in \hat{I}_k \tag{1.9}$$

instead of (1.4). Here $\hat{I}_k$ is a selected active index set, $\hat{c}_i(x_k)$ is a value associated with $c_i(x_k)$. Extensive numerical experiments in [20] showed that quadratic programming with fewer constraints in REQP would reduce computing time for computing the search direction. Actually, considering QP subproblems with a subset of the original constraints can be traced back to the work [19]. On the other hand, even for the classical SQP method, "warm starts" are used for the QP subproblems, which means it is always trying to solve the QP using the active set of the previous QP subproblem. Thus, an active set technique is used implicitly.

Another obvious advantage for using QP subproblems with few constraints is the reduction of the possibility of the inconsistence of the QP subproblems. It is well known that SQP may still fail in solving some well-defined inequality constrained problems with strict feasibility, since some of the QP subproblems may be inconsistent. Some research has been done for trying to circumvent the related difficulties, e.g., see [3,11,12].

Using active set technique in solving optimization with inequality constraints is not a new idea, for new references such as [9,10]. The attempt to improve the robustness of SQP has never stopped since its birth, there are a lot of references in the literature, the recent works include [12,18,24].

In this paper, we consider SQP in a more general framework. We present an active set SQP with general exact penalty functions (not necessarily the ℓ_1 exact penalty function). Our technique for selecting active constraints does not require the iteration points in the feasible region (see e.g. [15]). Our algorithm has some similarity to REQP, but it should be noted that by employing linearized inequality constraints (instead of equality constraints) the global convergence of our algorithm is established with weaker constraint qualification conditions. Under mild conditions, our algorithm can identify the active set of the original problem when the iterates are close to a solution. The superlinear convergence of our algorithm follows from the standard results of the SQP method (e.g., [6]).

This paper is organized as follows. In Section 2, we give the definition of the active set. Active constraints are selected based on the constraint violations as well as the dual variables. A new SQP algorithm applying our active techniques is given. The global convergence of the algorithm is proved in Section 3.

2. An active set SQP algorithm

Let x be the current iterate point and $\lambda = (\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(m)})^{\mathrm{T}} \in \Re^m$ be an approximate multiplier. Define $z = (x, \lambda)$. Let $\varepsilon \geqslant 0$ be a scalar. Define the ε-active set at x corresponding to λ:

$$I(z, \varepsilon) = \{i : c_i(x) \leqslant \lambda^{(i)} + \varepsilon\}. \tag{2.1}$$

The QP problem that we use as a subproblem is defined by Q(z, B):

$$\min \quad g(x)^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} B d \tag{2.2}$$

$$\text{s.t.} \quad c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I(z, \varepsilon). \tag{2.3}$$

Define the sets

$$S(z, \varepsilon) = \{d : c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I(z, \varepsilon)\}, \tag{2.4}$$

$$S_0(x) = \{d : c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I_0\}, \tag{2.5}$$

where $I_0 = \{1, 2, \ldots, m\}$. Then $S_0(x) \subset S(z, \varepsilon)$ since $I(z, \varepsilon) \subset I_0$.

Assume that there exists $v \in \Re^n$ such that

$$\nabla c_i(x)^{\mathrm{T}} v > 0, \quad i \in I(z, \varepsilon). \tag{2.6}$$

This condition is sufficient for $S(z, \varepsilon) \neq \emptyset$. When z is a Kuhn–Tucker point of (1.1)–(1.2) and $\varepsilon = 0$, (2.6) is precisely the *Mangasarian–Fromovitz constraint qualification* (MFCQ) at x. On the other hand, in order to keep (1.9) consistent, we need to assume that $\nabla c_i(x_k)\ (i \in \hat{I}_k)$ are linearly independent, which is stronger than (2.6) (see [23]).

Now suppose the sets (2.4) and (2.5) are both nonempty. If B is positive definite, the QP subproblems (1.3)–(1.4) and (2.2)–(2.3) have unique solutions d_0 and d_+, respectively. Assume that $\lambda_{|I(z,\varepsilon)|} = \{\lambda^{(i)} : i \in I(z, \varepsilon)\}$ is a multiplier vector corresponding to d_+. Let $\lambda^{(i)} = 0$ for $i \in I_0 \backslash I(z, \varepsilon)$ and $\lambda = \{\lambda^{(i)} : i \in I_0\}$, we call this λ the multiplier corresponding to (2.2)–(2.3). If $d_0 = d_+$, λ is also a multiplier corresponding to d_0. If $d_+ \in S_0(x)$, then we must have $d_0 = d_+$. Furthermore, we have the following result:

Lemma 2.1. *For any $z = (x, \lambda) \in \Re^{n+m}$, let $p(z, \varepsilon) = \min\{c_i(x)/\|\nabla c_i(x)\|_2 : i \in I_0 \backslash I(z, \varepsilon)\}$ and define the sets*

$$\hat{S}(z, \varepsilon) = \{d : c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I(z, \varepsilon) \text{ and } \|d\|_2 \leqslant p(z, \varepsilon)\}, \tag{2.7}$$

$$\hat{S}_0(z, \varepsilon) = \{d : c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I_0 \text{ and } \|d\|_2 \leqslant p(z, \varepsilon)\}. \tag{2.8}$$

Then we have $\hat{S}(z, \varepsilon) = \hat{S}_0(z, \varepsilon)$.

Proof. We only need to prove $\hat{S}(z, \varepsilon) \subset \hat{S}_0(z, \varepsilon)$.

For any $d \in \hat{S}(z, \varepsilon)$, (2.7) gives that $\|d\|_2 \leqslant c_i(x)/\|\nabla c_i(x)\|_2$ for any $i \in I_0 \backslash I(z, \varepsilon)$, i.e.,

$$c_i(x) - \|\nabla c_i(x)\|_2 \|d\|_2 \geqslant 0, \quad i \in I_0 \backslash I(z, \varepsilon). \tag{2.9}$$

Hence,

$$c_i(x) + \nabla c_i(x)^{\mathrm{T}} d \geqslant 0, \quad i \in I_0 \backslash I(z, \varepsilon), \tag{2.10}$$

which implies that $d \in \hat{S}_0(z, \varepsilon)$. This completes the proof. □

Define

$$\phi(x, \mu) = f(x) + \mu \psi(c(x)), \tag{2.11}$$

where $\psi(c(x)) = \operatorname{dist}(c(x)|\Re^m_+)$, $\Re^m_+ = \{x : x \geqslant 0, x \in \Re^m\}$ and

$$\operatorname{dist}(c(x)|\Re^m_+) = \inf\{\|c(x) - d\| : d \in \Re^m_+\}, \tag{2.12}$$

$\|\cdot\|$ is any given norm on $\Re^m$. Throughout the rest of the paper, if the norm is not specified, it is the same as that in (2.12).

It can be seen that x is a feasible point of (1.1)–(1.2) if and only if $\psi(c(x)) = 0$. If the norm is selected to be the ℓ_1 norm, then $\phi(x, \mu)$ is precisely the ℓ_1 exact penalty function.

Lemma 2.2. *Suppose that B is positive definite and d_+ is the unique solution of* (2.2)–(2.3). *There always exists a positive constant δ such that for $0 \leqslant \tau \leqslant \delta$,*

$$c_i(x) + \nabla c_i(x)^{\mathrm{T}} (\tau d_+) \geqslant 0, \quad i \in I_0 \backslash I(z, \varepsilon), \tag{2.13}$$

$$c_i(x) + \nabla c_i(x)^{\mathrm{T}} (\tau d_+) \geqslant (1 - \tau) c_i(x), \quad i \in I(z, \varepsilon). \tag{2.14}$$

Furthermore,

$$\psi(c(x)+\nabla c(x)^{\mathrm{T}}(\tau d_+))\leqslant(1-\tau)\psi(c(x)). \tag{2.15}$$

Proof. Let $\bar{I}(z,\varepsilon)=\{i\in I_0\backslash I(z,\varepsilon):\nabla c_i(x)^{\mathrm{T}}d_+<0\}$. We can see that (2.13) and (2.14) hold if we let

$$\delta=\min\left\{\min_{i\in\bar{I}(z,\varepsilon)}\frac{-c_i(x)}{\nabla c_i(x)^{\mathrm{T}}d_+};1\right\}. \tag{2.16}$$

Inequality (2.15) follows from (2.13)–(2.14). □

Let $d=\tau d_+$ in (2.13)–(2.14), then

$$c_i(x)+\nabla c_i(x)^{\mathrm{T}}d\geqslant 0,\quad i\in I_0\backslash I(z,\varepsilon), \tag{2.17}$$

$$\tau c_i(x)+\nabla c_i(x)^{\mathrm{T}}d\geqslant 0,\quad i\in I(z,\varepsilon). \tag{2.18}$$

The following result is known from [11,12].

Lemma 2.3. *Define $q(x)=\psi(c(x))$, for every d and x in $\Re^n$, the directional derivative $q'(x;d)$ exists and satisfies the inequality*

$$q'(x;d)\leqslant\psi(c(x)+\nabla c(x)^{\mathrm{T}}d)-\psi(c(x)). \tag{2.19}$$

Let d_+ be the unique solution of QP(z, B) and $d=\delta d_+$ satisfy (2.17)–(2.18). It follows from (2.15) that

$$\Delta(x)=\psi(c(x)+\nabla c(x)^{\mathrm{T}}d)-\psi(c(x))\leqslant-\delta\psi(c(x))\leqslant 0. \tag{2.20}$$

By Lemma 2.3, *d* is a descent direction of $q(x)$.

Lemma 2.4. *Suppose that $B\in\Re^{n\times n}$ is a symmetric positive definite matrix, d_+ is the unique solution of* (2.2)–(2.3), *$d=\delta d_+$ satisfies* (2.17)–(2.18). *Then*

(i) $\Delta(x)=0$ *if and only if x is a feasible point of* (1.1)–(1.2);
(ii) *Define $\theta(x,\mu)=g(x)^{\mathrm{T}}d+\mu\Delta(x)$. If x is a Kuhn–Tucker point of* (1.1)–(1.2), *then $\theta(x,\mu)=0$ for any $\mu\in\Re$*;
(iii) *If x is a feasible point but not a Kuhn–Tucker point of* (1.1)–(1.2), $\theta(x,\mu)<0$.

Proof. (i) If *x* is a feasible point of (1.1)–(1.2), $\psi(c(x))=0$. It follows from (2.20) that $\Delta(x)=0$. On the other hand, if $\Delta(x)=0$, (2.20) implies $\psi(c(x))=0$, which says that *x* is feasible.

(ii) If *x* is a Kuhn–Tucker point of (1.1)–(1.2), then *x* is feasible and $d_+=0$. Thus, $\theta(x,\mu)=0$ for any $\mu\in\Re$.

(iii) If *x* is a feasible point of (1.1)–(1.2), $\Delta(x)=0$ by (i). Since zero is feasible for the constraints of QP(z, B), $g(x)^{\mathrm{T}}d_+\leqslant g(x)^{\mathrm{T}}d_++\frac{1}{2}d_+^{\mathrm{T}}Bd_+\leqslant 0$. Moreover, $d_+\neq 0$, otherwise *x* is a Kuhn–Tucker point. By the fact that *B* is positive definite, we know that $g(x)^{\mathrm{T}}d<0$. Therefore, $\theta(x,\mu)<0$. □

The penalty parameter μ is chosen as follows. If *x* is not a feasible point of (1.1)–(1.2) and *d* is the same as that in Lemma 2.4, we set $\mu=\max\{(g(x)^{\mathrm{T}}d+d^{\mathrm{T}}Bd)/(-\Delta(x)),2\mu_-+\beta\}$. Otherwise, we let

$\mu = \mu_-$. Here, μ_- is the penalty parameter in the previous iteration and $\beta \geqslant 0$ is a constant. Consequently, Lemma 2.4 (ii) can be stated as: $\theta(x, \mu) = 0$ if and only if x is a Kuhn–Tucker point of (1.1)–(1.2).

The algorithm is described as follows.

Algorithm 2.5. Step 0: *Given* $x_0 \in \Re^n$, $\lambda_0 \in \Re^m$, $\lambda_0 \geqslant 0$, B_0 *is an* $n \times n$ *positive definite matrix*, μ_0, ε_0, ε *are positive scalars*, $\beta \geqslant 0$ *is a constant*, $0 < \sigma_1 \leqslant \sigma_2 < 1$, $\gamma_1 > 0$, $1 > \gamma_2 > 0$;

Step 1: *Solve* $\mathrm{QP}(z_k, B_k)$. *Suppose that* d'_k *is the solution. If* $\|d'_k\| \leqslant \varepsilon$, *then stop*;

Step 2: *Let* $\bar{I}_k = \{i \in I_0 \backslash I(z_k, \varepsilon_k) : \nabla c_i(x_k)^{\mathrm{T}} d'_k < 0\}$, *compute*

$$\delta_k = \min\left\{\min_{i \in \bar{I}_k} \frac{-c_i(x_k)}{\nabla c_i(x_k)^{\mathrm{T}} d'_k}; 1\right\} \tag{2.21}$$

and set $d_k = \delta_k d'_k$;

Step 3: *Compute* $\Delta(x_k)$. *If* $\Delta(x_k) = 0$ *or* $g(x_k)^{\mathrm{T}} d_k + \mu_k \Delta(x_k) \leqslant -d_k^{\mathrm{T}} B_k d_k$ *then set* $\mu_{k+1} = \mu_k$ *and goto* Step 4; *Otherwise, set*

$$\mu_{k+1} = \max\{(g(x_k)^{\mathrm{T}} d_k + d_k^{\mathrm{T}} B_k d_k)/(-\Delta(x_k)), 2\mu_k + \beta\}; \tag{2.22}$$

Step 4: *Let* $\theta(x_k, \mu_{k+1}) = g(x_k)^{\mathrm{T}} d_k + \mu_{k+1} \Delta(x_k)$ *and choose* α_k *such that*

$$\phi(x_k + \alpha_k d_k, \mu_{k+1}) \leqslant \phi(x_k, \mu_{k+1}) + \sigma_1 \alpha_k \theta(x_k, \mu_{k+1}) \tag{2.23}$$

and either $\alpha_k \geqslant \gamma_1$ *or there is an* $\bar{\alpha}_k > 0$ *such that* $\alpha_k \geqslant \gamma_2 \bar{\alpha}_k > 0$ *and*

$$\phi(x_k + \bar{\alpha}_k d_k, \mu_{k+1}) > \phi(x_k, \mu_{k+1}) + \sigma_2 \bar{\alpha}_k \theta(x_k, \mu_{k+1}); \tag{2.24}$$

Step 5: *Set* $x_{k+1} = x_k + \alpha_k d_k$, $\lambda^i_{k+1} = 0$ *for* $i \in I_0 \backslash I(z_k, \varepsilon_k)$; *Compute* B_{k+1} *by some quasi-Newton formulae*; *generate* ε_{k+1}; *set* $k = k + 1$ *and go to* Step 1.

Because the relation

$$\begin{aligned}\phi(x_k + \alpha d_k, \mu_{k+1}) - \phi(x_k, \mu_{k+1}) &= f(x_k + \alpha d_k) - f(x_k) + \mu_{k+1}(q(x_k + \alpha d_k) - q(x_k)) \\ &\leqslant \alpha g(x_k)^{\mathrm{T}} d_k + \alpha \mu_{k+1} \Delta(x_k) + \mathrm{o}(\alpha) \\ &= \alpha \theta(x_k, \mu_{k+1}) + \mathrm{o}(\alpha)\end{aligned}$$

holds for sufficiently small $\alpha > 0$, we can see that there exists $\alpha_k > 0$ satisfying the conditions in Step 4 of the algorithm.

Inequality (2.23) guarantees that $\phi(x_k + \alpha d_k, \mu_{k+1})$ decreases sufficiently and condition (2.24) keeps α_k away from zero. B_{k+1} can be computed by Powell's updating procedure [16] and [23].

3. Global convergence

To study the convergence properties of Algorithm 2.5, we need the following assumptions:

(A1) $f(x)$ and $c_i(x)$, $i \in I_0$ are twice continuously differentiable functions;

(A2) The approximation B_k of the Lagrangian Hessian is positive definite and there exists two positive constants m and M such that

$$M_1\|d\|_2^2 \leqslant d^{\mathrm{T}} B_k d \leqslant M_2\|d\|_2^2 \tag{3.1}$$

holds for all $d \in \Re^n$ and all $k \geqslant 1$.

First, we have the following lemma.

Lemma 3.1. *Suppose that* $\{x_k\}$ *generated by Algorithm* 2.5 *is contained in a closed convex set* Ω, $f(x)$, $g(x)$, $c(x)$, *and* $\nabla c(x)$ *are bounded on* Ω, (d_k', λ_{k+1}) *is a Kuhn–Tucker pair of* $\mathrm{QP}(z_k, \varepsilon_k)$.

(i) *If* $x_k \to x^*$ *for* $k \in K$, *where* x^* *is not necessarily a Kuhn–Tucker point of* (1.1)–(1.2), *and there exists* $v \in \mathbb{R}^n$ *such that*

$$\nabla c_i(x^*)^{\mathrm{T}} v > 0, \quad i \in \mathscr{I}(x^*, \varepsilon), \tag{3.2}$$

where $\mathscr{I}(x^*, \varepsilon) = \{i : i \in I(z_k, \varepsilon)$ *for infinitely many* $k \in K\}$ *and* ε *is a positive scalar, then* $\{\lambda_{k+1} : k \in K\}$ *is bounded.*

(ii) *For any infeasible point* x_k, $k \in K$, $[(g_k^{\mathrm{T}} d_k + d_k^{\mathrm{T}} B_k d_k)/-\Delta_k]$ *is bounded above.*

Proof. (i) First, we prove that $\|d_k'\|$ is bounded for $k \in K$. By (4.1), there is a $d^* \in \Re^n$ such that

$$c_i(x^*) + \nabla c_i(x^*)^{\mathrm{T}} d^* > 0, \quad i \in \mathscr{I}(x^*, \varepsilon). \tag{3.3}$$

Thus, there exists a $k_0 > 0$, for $k \geqslant k_0$,

$$c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d^* \geqslant 0, \quad i \in \mathscr{I}(x^*, \varepsilon). \tag{3.4}$$

It follows from the definition of $\mathscr{I}(x^*, \varepsilon)$ that there is a positive integer k_1 such that $I(z_k, \varepsilon_k) \subseteq \mathscr{I}(x^*, \varepsilon)$ for all $k \geqslant k_1$, $k \in K$. Thus, for all $k \geqslant \max\{k_0, k_1\}$, $k \in K$, d^* is a feasible point of $\mathrm{QP}(z_k, \varepsilon_k)$. It follows from (3.1) that for $k \in K$, $\|d_k'\|_2$ is bounded.

By Algorithm 2.5, for $i \notin \mathscr{I}(x^*, \varepsilon)$, $\lambda_{k+1}^{(i)} = 0$ holds except for a finite number of iterations. Now suppose that $\lambda_{k+1}^{(i)}$ $(i \in I(z_k, \varepsilon), k \in K)$ is unbounded. Let (d_k', λ_{k+1}) is the solution of $\mathrm{QP}(z_k, B_k)$, which implies

$$g_k + B_k d_k' - \sum_{i \in I(z_k, \varepsilon)} \lambda_{k+1}^{(i)} \nabla c_i(x_k) = 0 \tag{3.5}$$

for all $k \in K$. Normalizing by $\|\lambda_{k+1}\|_2$ and let $k \to \infty$, we have

$$\sum_{i \in \mathscr{I}(x^*, \varepsilon)} \bar{\lambda}^{(i)} \nabla c_i(x^*) = 0, \tag{3.6}$$

where $\bar{\lambda}^{(i)} \geqslant 0$ and $\|\bar{\lambda}\|_2 \neq 0$. (3.6) contradicts (3.2).

(ii) Since d_k' is the solution of $\mathrm{QP}(z_k, B_k)$, we have for $\lambda_{k+1} \geqslant 0$,

$$g(x_k)^{\mathrm{T}} d_k' + d_k'^{\mathrm{T}} B_k d_k' - \lambda_{k+1}^{\mathrm{T}} \nabla c(x_k)^{\mathrm{T}} d_k' = 0, \tag{3.7}$$

$$\lambda_{k+1}^{\mathrm{T}} (c(x_k) + \nabla c(x_k)^{\mathrm{T}} d_k') = 0. \tag{3.8}$$

Thus,

$$g(x_k)^{\mathrm{T}} d_k' + d_k'^{\mathrm{T}} B_k d_k' = -\lambda_{k+1}^{\mathrm{T}} c(x_k). \tag{3.9}$$

It follows from (3.9), $d_k = \delta_k d_k'$ and $\delta_k \in (0, 1]$ that

$$\begin{aligned}\frac{g_k^{\mathrm{T}} d_k + d_k^{\mathrm{T}} B_k d_k}{-\Delta_k} &= \frac{-\delta_k \lambda_{k+1}^{\mathrm{T}} c(x_k) - \delta_k(1-\delta_k) d_k'^{\mathrm{T}} B_k d_k'}{-\Delta_k} \\ &\leqslant \frac{-\delta_k \lambda_{k+1}^{\mathrm{T}} c(x_k)}{-\Delta_k} \leqslant \frac{-\lambda_{k+1}^{\mathrm{T}} c(x_k)}{\psi(c(x_k))}.\end{aligned} \tag{3.10}$$

Moreover, by the Cauchy–Schwarz inequality, we have $-\lambda_{k+1}^{\mathrm{T}} c(x_k) \leqslant \|\lambda_{k+1}\|_0 \psi(c(x_k))$, where $\|\cdot\|_0$ is the dual norm of $\|\cdot\|$. Thus, the result follows from (i) and (3.10). □

Lemma 3.2. *Under the assumptions of* (A1) *and* (A2), *suppose that* $f(x)$ *is bounded below on* R^n. *If the sequence* $\{x_k\}$ *is an infinite sequence generated by Algorithm* 2.5, $\{\mu_k\}$ *is bounded,* $\{x_k : k \in K\}$ *is any convergent subsequence, then* $d_k \to 0$ *for* $k \in K$ *and* $k \to \infty$.

Proof. First, we show that the inequality

$$\theta(x_k, \mu_{k+1}) \leqslant -\tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k \tag{3.11}$$

holds for all x_k. If x_k is not a feasible point of (1.1)–(1.2), (3.11) follows directly from the procedure of choosing the penalty parameter. Suppose that x_k is a feasible point of (1.1)–(1.2). Since $d = 0$ is feasible for QP(z_k,B_k), $g_k^{\mathrm{T}} d_k' + \frac{1}{2} d_k'^{\mathrm{T}} B_k d_k' \leqslant 0$. The definition of μ_{k+1} in Algorithm 2.5 implies that

$$\theta(x_k, \mu_{k+1}) = g_k^{\mathrm{T}} d_k \leqslant -\tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k. \tag{3.12}$$

Now we prove the limit of $\phi(x_k, \mu_k)$ exists for $k \to \infty$. The boundedness of μ_k indicates that there exists k_0 such that $\mu_k = \mu_{k_0} = \bar{\mu}$ for all $k \geqslant k_0$. Therefore, it follows from (2.23) that

$$\phi(x_{k+1}, \mu_{k+1}) < \phi(x_k, \mu_{k+1}) = \phi(x_k, \mu_k) \tag{3.13}$$

for all $k \geqslant k_0$. Moreover, $\phi(x_k, \mu_k) \geqslant f(x_k)$. Thus, $\lim_{k\to\infty} \phi(x_k, \mu_k)$ exists. Again by (2.23), for $k \geqslant k_0$,

$$\begin{aligned}\sigma_1 \alpha_k \theta(x_k, \mu_{k+1}) &\geqslant \phi(x_{k+1}, \mu_{k+1}) - \phi(x_k, \mu_{k+1}) \\ &= \phi(x_{k+1}, \mu_{k_0}) - \phi(x_k, \mu_{k_0}).\end{aligned} \tag{3.14}$$

It follows from Step 3 of Algorithm 2.5 and Lemma 2.4 that $\theta(x_k, \mu_{k+1}) \leqslant 0$ for all k, so we have

$$\alpha_k \theta(x_k, \mu_{k+1}) \to 0 \tag{3.15}$$

for $k \to \infty$.

Now, we complete our proof by contradiction. If the lemma is not true, there is a positive constant η such that $\bar{K} = \{k : \|d_k\|_2 > \eta, k \in K\}$ is an infinite set. It follows from (3.11) and (3.1) that

$$\theta(x_k, \mu_{k+1}) \leqslant -\tfrac{1}{2} M_1 \eta^2 < 0, \quad \forall k \in \bar{K}. \tag{3.16}$$

Hence, (3.15) implies that $\lim_{k\to\infty, k\in\bar{K}} \alpha_k = 0$. Therefore for sufficiently large $k \in \bar{K}$, (2.24) holds. We may assume $\bar{\alpha}_k < 1$ for all $k \in \bar{K}$ by Step 4. Observe that

$$c(x_k + \bar{\alpha}_k d_k) - (c(x_k) + \bar{\alpha}_k \nabla c(x_k)^{\mathrm{T}} d_k) = \int_0^1 (\nabla c(x_k + t\bar{\alpha}_k d_k) - \nabla c(x_k))^{\mathrm{T}}\, \mathrm{d}t \bar{\alpha}_k\, d_k$$
$$= [m_1(\bar{\alpha}_k d_k)]^{\mathrm{T}} (\bar{\alpha}_k d_k), \tag{3.17}$$

$$f(x_k + \bar{\alpha}_k d_k) - (f(x_k) + \bar{\alpha}_k g(x_k)^{\mathrm{T}} d_k) = \int_0^1 (g(x_k + t\bar{\alpha}_k d_k) - g(x_k))^{\mathrm{T}}\, \mathrm{d}t \bar{\alpha}_k d_k$$
$$= [m_2(\bar{\alpha}_k d_k)]^{\mathrm{T}} (\bar{\alpha}_k d_k), \tag{3.18}$$

where

$$m_1(\bar{\alpha}_k d_k) = \int_0^1 (\nabla c(x_k + t\bar{\alpha}_k d_k) - \nabla c(x_k))^{\mathrm{T}}\, \mathrm{d}t, \tag{3.19}$$

$$m_2(\bar{\alpha}_k d_k) = \int_0^1 (g(x_k + t\bar{\alpha}_k d_k) - g(x_k))^{\mathrm{T}}\, \mathrm{d}t. \tag{3.20}$$

Therefore, it follows the above relations and (2.24)–(2.25) that

$$\begin{aligned} q(x_k + \bar{\alpha}_k d_k) - q(x_k) &\leqslant \|c(x_k + \bar{\alpha}_k d_k) - (c(x_k) + \bar{\alpha}_k \nabla c(x_k)^{\mathrm{T}} d_k)\| \\ &\quad + \psi(c(x_k) + \bar{\alpha}_k \nabla c(x_k)^{\mathrm{T}} d_k) - \psi(c(x_k)) \\ &\leqslant \bar{\alpha}_k \|d_k\| \|m_1(\bar{\alpha}_k d_k)\| + \bar{\alpha}_k [\psi(c(x_k) + \nabla c(x_k)^{\mathrm{T}} d_k) - \psi(c(x_k))] \\ &\leqslant c_1 \bar{\alpha}_k \|d_k\|_2 \|m_1(\bar{\alpha}_k d_k)\| + \bar{\alpha}_k \Delta(x_k) \end{aligned} \tag{3.21}$$

and

$$\sigma_2 \bar{\alpha}_k \theta(x_k, \bar{\mu}) < \phi(x_k + \bar{\alpha}_k d_k, \bar{\mu}) - \phi(x_k, \bar{\mu}) \tag{3.22}$$
$$\leqslant \bar{\alpha}_k \theta(x_k, \bar{\mu}) + \bar{\alpha}_k \|d_k\|_2 (\bar{\mu} c_1 \|m_1(\bar{\alpha}_k d_k)\| + c_2 \|m_2(\bar{\alpha}_k d_k)\|), \tag{3.23}$$

where c_1 and c_2 are positive scalars. The above two inequalities and (3.16) give

$$\begin{aligned} \tfrac{1}{2}(1-\sigma_2) M_1 \|d_k\|_2^2 &\leqslant -(1-\sigma_2)\theta(x_k, \bar{\mu}) \\ &< \|d_k\|_2 (\bar{\mu} c_1 \|m_1(\bar{\alpha}_k d_k)\| + c_2 \|m_2(\bar{\alpha}_k d_k)\|). \end{aligned} \tag{3.24}$$

Divided by $\|d_k\|_2$, the above relation implies, when $k \in \bar{K}$ and $k \to \infty$,

$$(1-\sigma_2) M_1 \leqslant 0, \tag{3.25}$$

which is a contradiction. This completes our proof. □

Corollary 3.3. *Under the assumptions of Lemma* 3.2, *for sufficiently large* $k \in K$, d_k *is identical to the direction generated by* (1.3)–(1.4).

Proof. The result follows from Lemma 3.2, Algorithm 2.5 and Lemma 2.1. □

Theorem 3.4. *Under the assumptions of* (A1) *and* (A2), *suppose that* $f(x)$ *is bounded below on* $\Re^n$. *If the sequence* $\{x_k\}$ *is an infinite sequence generated by Algorithm* 2.5, $\{\mu_k\}$ *is bounded, then any accumulation point of* $\{x_k\}$ *is a Kuhn–Tucker point of* (1.1)–(1.2).

Proof. Without loss of generality, we may assume $\mu_k = \mu$ for all k. Let x^* be an accumulation point of $\{x_k\}$, then there exists a subset K such that $x_k \to x^*$ $(k \to \infty, k \in K)$. If $\theta(x^*, \mu) = 0$, then by Lemma 2.4, x^* is a Kuhn–Tucker point of (1.1)–(1.2).

Now we assume $\theta(x^*, \mu) < 0$. Then there exists $\hat{k}$ such that

$$\theta(x_k, \mu) \leqslant \tfrac{1}{2}\theta(x^*, \mu) < 0 \tag{3.26}$$

for all $k \geqslant \hat{k}$, $k \in K$. From the proof of Lemma 3.2, we have

$$d_k \to 0 \quad (k \to \infty, k \in K). \tag{3.27}$$

(3.27) and the definition of $\theta(x_k, \mu)$ give that

$$\theta(x_k, \mu) \to 0 \quad (k \to \infty, k \in K), \tag{3.28}$$

which contradicts (3.26). $\square$

The following result shows that our algorithm can correctly identify the active constraints and solve the QP subproblem with only active constraints automatically, provided that $\varepsilon_k \to 0$ and the iterate is sufficiently close to the solution.

Lemma 3.5. *Under the assumptions of Theorem* 3.4, *suppose that* $x_k \to x^*(k \to \infty)$, $\nabla c_i(x^*)$ $(i \in I^*(x^*))$ *are linearly independent,* λ^* *is the multiplier associated with* x^*. *If* $\varepsilon_k \to 0$, *the strict complementarity condition holds at* z^*, *then* $I(z_k, \varepsilon_k) = I^*(z^*)$ *for all sufficiently large k.*

Proof. The assumption implies that $\lambda_k \to \lambda^*$ for $k \to \infty$. Thus, $z_k \to z^*$. It follows from (2.1) that

$$I(z^*, 0) = \{i : c_i(x^*) \leqslant \lambda^{*(i)}\}. \tag{3.29}$$

Let

$$I_1(z^*, 0) = \{i : c_i(x^*) < \lambda^{*(i)}\}, \quad I_2(z^*, 0) = \{i : c_i(x^*) = \lambda^{*(i)}\}, \tag{3.30}$$

then by the strict complementarity condition at z^*, we have

$$I_1(z^*, 0) = I^*(z^*) \quad \text{and} \quad I_2(z^*, 0) = \emptyset. \tag{3.31}$$

Similarly, define

$$I_1(z_k, \varepsilon_k) = \{i : c_i(x_k) < \lambda_k^{(i)} + \varepsilon_k\}, \quad I_2(z_k, \varepsilon_k) = \{i : c_i(x_k) = \lambda_k^{(i)} + \varepsilon_k\}. \tag{3.32}$$

It is straightforward that there exists a positive integer k_1 such that for $k \geqslant k_1$,

$$I_1(z^*, 0) \subseteq I_1(z_k, \varepsilon_k) \quad \text{and} \quad I_2(z^*, 0) \supseteq I_2(z_k, \varepsilon_k). \tag{3.33}$$

Hence,

$$I_2(z_k, \varepsilon_k) = \emptyset \quad \text{and} \quad I(z^*, 0) \subseteq I(z_k, \varepsilon_k) \quad \text{for } k \geqslant k_1. \tag{3.34}$$

It is easy to see that there exists k_2 such that $k \geqslant k_2$,

$$I(z^*, 0) \supseteq I(z_k, \varepsilon_k). \tag{3.35}$$

Therefore, (3.34) and (3.35) give that

$$I(z^*, 0) = I(z_k, \varepsilon_k) \tag{3.36}$$

for $k \geqslant \max\{k_1, k_2\}$, which completes the proof. □

References

[1] D.P. Bertsekas, Constrained Optimization and Lagrange Multipliers, Academic Press, New York, 1982.
[2] M.C. Biggs, Constrained minimization using recursive equality quadratic programming, in: F.A. Lootsma (Ed.), Numerical Methods for Nonlinear Optimization, Academic Press, New York, 1972, pp. 411–428.
[3] J.V. Burke, S.P. Han, A robust sequential quadratic programming method, Math. Programming 43 (1989) 277–303.
[4] R. Fletcher, Practical Methods of Optimization, Second ed., Wiley, Chichester, 1987.
[5] R. Fletcher, S. Leyffer, D. Ralph, S. Scholtes, Local convergence of SQP methods for mathematical programs with equilibrium constraints, Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, 2002.
[6] Ph.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, New York, 1981.
[7] S.P. Han, A globally convergent method for nonlinear programming, J. Optim. Theory. Appl. 22 (1977) 297–309.
[8] S.P. Han, J.S. Pang, N. Rangaraj, Globally convergent Newton methods for nonsmooth equations, Math. Oper. Res. 17 (1992) 586–607.
[9] C. Kanzow, An active set-type Newton method for constrained nonlinear systems, Complementarity: Applications, Algorithms and Extensions (Madison, WI, 1999), Applications in Optimizations, vol. 50, Kluwer Academic Publishers, Dordrecht, 2001, pp. 179–200.
[10] K. Kunisch, A. Rösch, Primal–dual active set strategy for a general class of constrained optimal control problems, SIAM J. Optim. 13 (2002) 321–334.
[11] X.-W. Liu, Y. Yuan, Global convergence of an active set SQP algorithm, Research Report ICM-97-82, ICMSEC.
[12] X.-W. Liu, Y.-X. Yuan, A robust algorithm for optimization with general equality and inequality constraints, SIAM J. Sci. Comput. 22 (2000) 517–534.
[13] X.-W. Liu, G. Zhao, A decomposition method based on SQP for a class of multistage stochastic programs, SIAM J. Optim. 14 (2003) 200–222.
[14] W. Murray, F.J. Prieto, A sequential quadratic programming algorithm using an incomplete solution of the subproblem, SIAM J. Optim. 5 (1995) 590–640.
[15] J.S. Pang, Serial and parallel computation of Karush–Kuhn–Tucker points via nonsmooth equations, SIAM J. Optim. 4 (1994) 872–893.
[16] M.J.D. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: G.A. Watson (Ed.), Proceedings of the 1977 Dundee Biennial Conference on Numerical Analysis, Springer, Berlin, 1978, pp. 144–157.
[17] M.J.D. Powell, The convergence of variable metric methods for nonlinear constrained optimization calculations, in: O.L. Mangasarian, R.R. Meyer, S.M. Robinson (Eds.), Nonlinear Programming 3, Academic Press, New York, 1978, pp. 27–63.
[18] M.J.D. Powell, Y. Yuan, A recursive quadratic programming algorithm that using differentiable exact penalty function, Math. Programming 35 (1986) 265–278.
[19] B.N. Pshenichny, Danilin, Numerical Methods in Extremal Problems, Mir Publishers, Moscow, 1978.
[20] K. Schittkowski, NLPQL: a FORTRAN subroutine solving constrained nonlinear programming problems, Ann. Oper. Res. 5 (1986) 485–500.
[21] J. Stoer, Principles of sequential quadratic programming methods for solving nonlinear programs, in: K. Schittkowski (Ed.), NATO ASI Series, vol. F15, Computational Mathematical Programming, Springer, Berlin.
[22] K. Taji, M. Fukushima, A new merit function and a successive quadratic programming algorithm for variational inequality problems, SIAM J. Optim. 6 (1996) 704–713.
[23] Y. Yuan, Numerical Methods for Nonlinear Programming, Modern Mathematics Series, Shanghai Scientific & Technical Publishers, 1993.
[24] J.L. Zhang, X.S. Zhang, A robust SQP method for optimization with inequality constraints, J. Comput. Math. 21 (2003) 247–256.